

# Confidentiality-Preserving Data Publishing for Credulous Users by Extended Abduction

Katsumi Inoue<sup>1</sup>, Chiaki Sakama<sup>2</sup> and Lena Wiese<sup>1\*</sup>

<sup>1</sup> National Institute of Informatics  
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan  
{ki|wiese}@nii.ac.jp

<sup>2</sup> Department of Computer and Communication Sciences Wakayama University  
930 Sakaedani, Wakayama 640-8510, Japan  
sakama@sys.wakayama-u.ac.jp

**Abstract.** Publishing private data on external servers incurs the problem of how to avoid unwanted disclosure of confidential data. We study a problem of confidentiality in extended disjunctive logic programs and show how it can be solved by extended abduction. In particular, we analyze how credulous non-monotonic reasoning affects confidentiality.

**Keywords:** Data publishing, confidentiality, privacy, extended abduction, answer set programming, negation as failure, non-monotonic reasoning

## 1 Introduction

Confidentiality of data (also called privacy or secrecy in some contexts) is a major security goal. Releasing data to a querying user without disclosing confidential information has long been investigated in areas like access control,  $k$ -anonymity, inference control, and data fragmentation. Such approaches prevent disclosure according to some security policy by restricting data access (denial, refusal), by modifying some data (perturbation, noise addition, cover stories, lying, weakening), or by breaking sensitive associations (fragmentation). Several approaches (like [3,8,13,14,2,15]) employ logic-based mechanisms to ensure data confidentiality. In particular, [5] use brave reasoning in default logic theories to solve a privacy problem in a classical database (a set of ground facts). For a non-classical knowledge base (where negation as failure *not* is allowed) [16] study correctness of access rights. Confidentiality of predicates in collaborative multi-agent abduction is a topic in [10].

In this article we analyze **confidentiality-preserving data publishing** in a knowledge base setting: data as well as integrity constraints or deduction rules are represented as logical formulas. If such a knowledge base is released to the public for general querying (e.g., microcensus data) or outsourced to a storage provider (e.g., database-as-a-service in cloud computing), confidential data could be disclosed. We assume that users accessing the published knowledge base use a form of credulous (also called brave) reasoning to retrieve data from it; users also possess some invariant “a priori knowledge” that can be applied to these data to deduce further information. On the knowledge base side, a confidentiality policy specifies which is the confidential information that must never be disclosed. This paper is one of only few papers (see [11,16,10]) covering confidentiality for logic programs. This formalism however has relevance in multi-agent communications where agent knowledge is modeled by

---

\* Lena Wiese gratefully acknowledges a postdoctoral research grant of the German Academic Exchange Service (DAAD).

logic programs. With **extended abduction** ([12]) we obtain a “secure version” of the knowledge base that can safely be published even when a priori knowledge is applied. We show that computing the secure version for a credulous user corresponds to finding a skeptical anti-explanation for all the elements of the confidentiality policy. Extended abduction has been used in different applications like for example providing a logical framework for dishonest reasoning [11]. It can be solved by computing the answer sets of an update program (see [12]); thus an implementation of extended abduction can profit from current answer set programming (ASP) solvers [4]. To retrieve the confidentiality-preserving knowledge base  $K^{pub}$  from the input knowledge base  $K$ , the a priori knowledge *prior* and the confidentiality policy *policy*, a row of transformations are applied; the overall approach is depicted in Figure 1.

In sum, this paper makes the following contributions:

- it formalizes confidentiality-preserving data publishing for a user who retrieves data under a credulous query response semantics.
- it devises a procedure to securely publish a logic program (with an expressiveness up to extended disjunctive logic programs) respecting a subset-minimal change semantics.
- it shows that confidentiality-preservation for credulous users corresponds to finding a skeptical anti-explanation and can be solved by extended abduction.

In the remainder of this article, Section 2 provides background on extended disjunctive logic programs and answer set semantics; Section 3 defines the problem of confidentiality in data publishing; Section 4 recalls extended abduction and update programs; Section 5 shows how answer sets of update programs correspond to confidentiality-preserving knowledge bases; and Section 6 gives some discussion and concluding remarks.

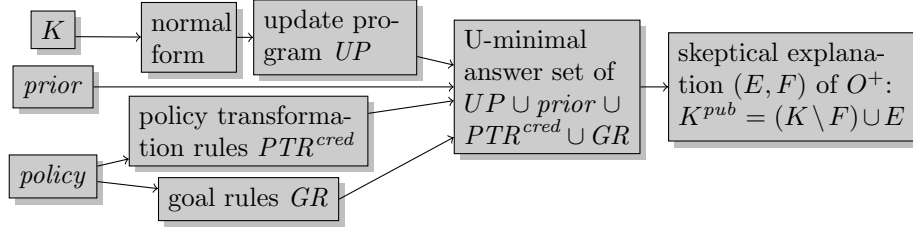
## 2 EDPs and answer set semantics

In this article, a knowledge base  $K$  is represented by an *extended disjunctive logic program* (EDP) – a set of formulas called *rules* of the form:

$$L_1; \dots; L_l \leftarrow L_{l+1}, \dots, L_m, \text{not} L_{m+1}, \dots, \text{not} L_n \quad (n \geq m \geq l \geq 0)$$

A rule contains literals  $L_i$ , disjunction “;”, conjunction “,”, negation as failure “*not*”, and material implication “ $\leftarrow$ ”. A literal is a first-order atom or an atom preceded by classical negation “ $\neg$ ”. *not* $L$  is called a *NAF-literal*. The disjunction left of the implication  $\leftarrow$  is called the *head*, while the conjunction right of  $\leftarrow$  is called the *body* of the rule. For a rule  $R$ , we write  $\text{head}(R)$  to denote the set of literals  $\{L_1, \dots, L_l\}$  and  $\text{body}(R)$  to denote the set of (NAF-)literals  $\{L_{l+1}, \dots, L_m, \text{not} L_{m+1}, \dots, \text{not} L_n\}$ . Rules consisting only of a singleton head  $L \leftarrow$  are identified with the literal  $L$  and used interchangeably. An EDP is ground if it contains no variables. If an EDP contains variables, it is identified with the set of its ground instantiations: the elements of its Herbrand universe are substituted in for the variables in all possible ways. We assume that the language contains no function symbol, so that each rule with variables represents a finite set of ground rules. For a program  $K$ , we denote  $\mathcal{L}_K$  the set of ground literals in the language of  $K$ . Note that EDPs offer a high expressiveness including disjunctive and non-monotonic reasoning.

*Example 1.* In a medical knowledge base  $\text{Ill}(x, y)$  states that a patient  $x$  is ill with disease  $y$ ;  $\text{Treat}(x, y)$  states that  $x$  is treated with medicine  $y$ . Assume that if you read the record and



**Fig. 1.** Finding a confidentiality-preserving  $K^{pub}$  for a credulous user

find that one treatment (Medi1) is recorded and another one (Medi2) is not recorded, then you know that the patient is at least ill with Aids or Flu (and possibly has other illnesses).

$K = \{Ill(x, Aids); Ill(x, Flu) \leftarrow Treat(x, Medi1), not\ Treat(x, Medi2) ,$   
 $Ill(Mary, Aids) , Treat(Pete, Medi1)\}$  serves as a running example.

The semantics of  $K$  can be given by the answer set semantics [7]: A set  $S \subseteq \mathcal{L}_K$  of ground literals *satisfies* a ground literal  $L$  if  $L \in S$ ;  $S$  satisfies a conjunction if it satisfies every conjunct;  $S$  satisfies a disjunction if it satisfies at least one disjunct;  $S$  satisfies a ground rule if whenever the body literals are contained in  $S$  ( $\{L_{l+1}, \dots, L_m\} \subseteq S$ ) and all NAF-literals are not contained in  $S$  ( $\{L_{m+1}, \dots, L_n\} \cap S = \emptyset$ ), then at least one head literal is contained in  $S$  ( $L_i \in S$  for an  $i$  such that  $1 \leq i \leq l$ ). If an EDP  $K$  contains no NAF-literals ( $m = n$ ), then such a set  $S$  is an *answer set* of  $K$  if  $S$  is a subset-minimal set such that

1.  $S$  satisfies every rule from the ground instantiation of  $K$ ,
2. If  $S$  contains a pair of complementary literals  $L$  and  $\neg L$ , then  $S = \mathcal{L}_K$ .

This definition of an answer set can be extended to full EDPs (containing NAF-literals) as in [12]: For an EDP  $K$  and a set of ground literals  $S \subseteq \mathcal{L}_K$ ,  $K$  can be transformed into a NAF-free program  $K^S$  as follows. For every ground rule from the ground instantiation of  $K$  (with respect to its Herbrand universe), the rule  $L_1; \dots; L_l \leftarrow L_{l+1}, \dots, L_m$  is in  $K^S$  if  $\{L_{m+1}, \dots, L_n\} \cap S = \emptyset$ . Then,  $S$  is an answer set of  $K$  if  $S$  is an answer set of  $K^S$ . An answer set is *consistent* if it is not  $\mathcal{L}_K$ . A program  $K$  is *consistent* if it has a consistent answer set; otherwise  $K$  is *inconsistent*.

*Example 2.* The example  $K$  has the following two consistent answer sets

$$S_1 = \{Ill(Mary, Aids), Treat(Pete, Medi1), Ill(Pete, Aids)\}$$

$$S_2 = \{Ill(Mary, Aids), Treat(Pete, Medi1), Ill(Pete, Flu)\}$$

When adding the negative fact  $\neg Ill(Pete, Flu)$  to  $K$ , then there is just one consistent answer set left: for  $K' := K \cup \{\neg Ill(Pete, Flu)\}$  the unique answer set is

$$S' = \{Ill(Mary, Aids), \neg Ill(Pete, Flu), Treat(Pete, Medi1), Ill(Pete, Aids)\}.$$

If a rule  $R$  is satisfied in *every* answer set of  $K$ , we write  $K \models R$ . In particular,  $K \models L$  if a literal  $L$  is included in every answer set of  $K$ .

### 3 Confidentiality-Preserving Knowledge Bases

When publishing a knowledge base  $K$  while preserving confidentiality of some data in  $K$  we do this according to

- the query response semantics that a user querying the published knowledge base applies; we focus on credulous query response semantics
- a confidentiality policy (denoted *policy*) describing confidential information that should not be released to the public
- background (a priori) knowledge (denoted *prior*) that a user can combine with query responses from the published knowledge base

First we define the credulous query response semantics: a ground formula  $Q$  is *true* in  $K$ , if  $Q$  is satisfied in some answer set of  $K$  – that is, there might be answer sets that do not satisfy  $Q$ . If a rule  $Q$  is non-ground and contains some free variables, the credulous response of  $K$  is the set of ground instantiations of  $Q$  that are *true* in  $K$ .

**Definition 1 (Credulous query response semantics).** *Let  $U$  be the Herbrand universe of a consistent knowledge base  $K$ . The credulous query responses of formula  $Q(X)$  (with a vector  $X$  of free variables) in  $K$  are*

$$\text{cred}(K, Q(X)) = \{Q(A) \mid A \text{ is a vector of elements } a \in U \text{ and there is an answer set of } K \text{ that satisfies } Q(A)\}$$

In particular, for a ground formula  $Q$ ,

$$\text{cred}(K, Q) = \begin{cases} Q & \text{if } K \text{ has an answer set that satisfies } Q \\ \emptyset & \text{otherwise} \end{cases}$$

It is usually assumed that in addition to the query responses a user has some additional knowledge that he can apply to the query responses. Hence, we additionally assume given a set of rules as some *invariant a priori knowledge prior*. Without loss of generality we assume that *prior* is an EDP. Thus, the priori knowledge may consist of additional facts that the user assumes to hold in  $K$ , or some rules that the user can apply to data in  $K$  to deduce new information.

A **confidentiality policy** *policy* specifies confidential information. We assume that *policy* contains only conjunctions of (NAF-)literals. However, see Section 5.1 for a brief discussion on how to use more expressive policy formulas. We do not only have to avoid that the published knowledge base contains confidential information but also prevent the user from deducing confidential information with the help of his a priori knowledge; this is known as the inference problem [6,2].

*Example 3.* If we wish to declare the disease aids as confidential for any patient  $x$  we can do this with *policy* =  $\{Ill(x, Aids)\}$ . A user querying  $K^{pub}$  might know that a person suffering from flu is not able to work. Hence *prior* =  $\{\neg AbleToWork(x) \leftarrow Ill(x, Flu)\}$ . If we wish to also declare a lack of work ability as confidential, we can add this to the confidentiality policy: *policy'* =  $\{Ill(x, Aids), \neg AbleToWork(x)\}$ .

Next, we establish a definition of confidentiality-preservation that allows for the answer set semantics as an inference mechanism and respects the credulous query response semantics: when treating elements of the confidentiality policy as queries, the credulous responses must be empty.

**Definition 2 (Confidentiality-preservation for credulous user).** A knowledge base  $K^{pub}$  preserves confidentiality of a given confidentiality policy under the credulous query response semantics and with respect to a given a priori knowledge prior, if for every conjunction  $C(X)$  in the policy, the credulous query responses of  $C(X)$  in  $K^{pub} \cup \text{prior}$  are empty:  $\text{cred}(K^{pub} \cup \text{prior}, C(X)) = \emptyset$ .

Note that in this definition the Herbrand universe of  $K^{pub} \cup \text{prior}$  is applied in the query response semantics; hence, free variables in policy elements  $C(X)$  are instantiated according to this universe. Note also that  $K^{pub} \cup \text{prior}$  must be consistent. Confidentiality-preservation for *skeptical* query response semantics is topic of future work.

A goal secondary to confidentiality-preservation is minimal change: We want to publish as many data as possible and want to modify these data as little as possible. Different notions of minimal change are used in the literature (see for example [1] for a collection of minimal change semantics in a data integration setting). We apply a subset-minimal change semantics: we choose a  $K^{pub}$  that differs from  $K$  only subset-minimally. In other words, there is not other confidentiality-preserving knowledge base  $K^{pub'}$  which inserts (or deletes) less rules to (from)  $K$  than  $K^{pub}$ .

**Definition 3 (Subset-minimal change).** A confidentiality-preserving knowledge base  $K^{pub}$  subset-minimally changes  $K$  (or is minimal, for short) if there is no confidentiality-preserving knowledge base  $K^{pub'}$  such that  $((K \setminus K^{pub'}) \cup (K^{pub'} \setminus K)) \subset ((K \setminus K^{pub}) \cup (K^{pub} \setminus K))$ .

*Example 4.* For the example  $K$  and policy and no a priori knowledge, the fact  $\text{Ill}(\text{Mary}, \text{Aids})$  has to be deleted. But also  $\text{Ill}(\text{Pete}, \text{Aids})$  can be deduced credulously, because it is satisfied by answer set  $S_1$ . In order to avoid this, we have three options: delete  $\text{Treat}(\text{Pete}, \text{Medi1})$ , delete the non-literal rule in  $K$  or insert  $\text{Treat}(\text{Pete}, \text{Medi2})$ . The same solutions are found for  $K$ ,  $\text{policy}'$  and  $\text{prior}$ : they block the credulous deduction of  $\neg \text{AbleToWork}(\text{Pete})$ . The same applies to  $K'$  and  $\text{policy}$ .

In the following sections we obtain a minimal solution  $K^{pub}$  for a given input  $K$ ,  $\text{prior}$  and  $\text{policy}$  by transforming the input into a problem of *extended abduction* and solving it with an appropriate update program.

## 4 Extended Abduction

Traditionally, given a knowledge base  $K$  and an observation formula  $O$ , *abduction* finds a “(positive) explanation”  $E$  – a set of hypothesis formulas – such that every answer set of the knowledge base and the explanation together satisfy the observation; that is,  $K \cup E \models O$ . Going beyond that [9,12] use *extended abduction* with the notions of “negative observations”, “negative explanations”  $F$  and “anti-explanations”. An abduction problem in general can be restricted by specifying a designated set  $\mathcal{A}$  of *abducibles*. This set poses syntactical restrictions on the explanation sets  $E$  and  $F$ . In particular, positive explanations are characterized by  $E \subseteq \mathcal{A} \setminus K$  and negative explanations by  $F \subseteq K \cap \mathcal{A}$ . If  $\mathcal{A}$  contains a formula with variables, it is meant as a shorthand for all ground instantiations of the formula. In this sense, an EDP  $K$  accompanied by an EDP  $\mathcal{A}$  is called an *abductive program* written as  $\langle K, \mathcal{A} \rangle$ . The aim of extended abduction is then to find (anti-)explanations as follows (where in this article only *skeptical* (anti-)explanations are needed):

- given a *positive* observation  $O$ , find a pair  $(E, F)$  where  $E$  is a positive explanation and  $F$  is a negative explanation such that
  1. [**skeptical explanation**]  $O$  is satisfied in *every* answer set of  $(K \setminus F) \cup E$ ; that is,  $(K \setminus F) \cup E \models O$
  2. [**consistency**]  $(K \setminus F) \cup E$  is consistent
  3. [**abducibility**]  $E \subseteq \mathcal{A} \setminus K$  and  $F \subseteq \mathcal{A} \cap K$
- given a *negative* observation  $O$ , find a pair  $(E, F)$  where  $E$  is a positive anti-explanation and  $F$  is a negative anti-explanation such that
  1. [**skeptical anti-explanation**] there is *no* answer set of  $(K \setminus F) \cup E$  in which  $O$  is satisfied
  2. [**consistency**]  $(K \setminus F) \cup E$  is consistent
  3. [**abducibility**]  $E \subseteq \mathcal{A} \setminus K$  and  $F \subseteq \mathcal{A} \cap K$

Among (anti-)explanations, **minimal** (anti-)explanations characterize a subset-minimal alteration of the program  $K$ : an (anti-)explanation  $(E, F)$  of an observation  $O$  is called minimal if for any (anti-)explanation  $(E', F')$  of  $O$ ,  $E' \subseteq E$  and  $F' \subseteq F$  imply  $E' = E$  and  $F' = F$ .

For an abductive program  $\langle K, \mathcal{A} \rangle$  both  $K$  and  $\mathcal{A}$  are semantically identified with their ground instantiations with respect to the Herbrand universe, so that set operations over them are defined on the ground instances. Thus, when  $(E, F)$  contain formulas with variables,  $(K \setminus F) \cup E$  means deleting every instance of formulas in  $F$ , and inserting any instance of formulas in  $E$  from/into  $K$ . When  $E$  contains formulas with variables, the set inclusion  $E' \subseteq E$  is defined for any set  $E'$  of instances of formulas in  $E$ . Generally, given sets  $S$  and  $T$  of literals/rules containing variables, any set operation  $\circ$  is defined as  $S \circ T = \text{inst}(S) \circ \text{inst}(T)$  where  $\text{inst}(S)$  is the ground instantiation of  $S$ . For example, when  $p(x) \in T$ , for any constant  $a$  occurring in  $T$ , it holds that  $\{p(a)\} \subseteq T$ ,  $\{p(a)\} \setminus T = \emptyset$ , and  $T \setminus \{p(a)\} = (T \setminus \{p(x)\}) \cup \{p(y) \mid y \neq a\}$ , etc. Moreover, any literal/rule in a set is identified with its variants modulo variable renaming.

#### 4.1 Normal form

Although extended abduction can handle the very general format of EDPs, some syntactic transformations are helpful. Based on [12] we will briefly describe how a semantically equivalent normal form of an abductive program  $\langle K, \mathcal{A} \rangle$  is obtained – where both the program  $K$  and the set  $\mathcal{A}$  of abducibles are EDPs. This makes an automatic handling of abductive programs easier; for example, abductive programs in normal form can be easily transformed into update programs as described in Section 4.2. The main step is that rules in  $\mathcal{A}$  can be mapped to atoms by a naming function  $n$ . Let  $\mathcal{R}$  be the set of abducible rules:

$$\mathcal{R} = \{\Sigma \leftarrow \Gamma \mid (\Sigma \leftarrow \Gamma) \in \mathcal{A} \text{ and } (\Sigma \leftarrow \Gamma) \text{ is not a literal}\}$$

Then the *normal form*  $\langle K^n, \mathcal{A}^n \rangle$  is defined as follows where  $n(R)$  maps each rule  $R$  to a fresh atom with the same free variables as  $R$ :

$$\begin{aligned} K^n &= (K \setminus \mathcal{R}) \cup \{\Sigma \leftarrow \Gamma, n(R) \mid R = (\Sigma \leftarrow \Gamma) \in \mathcal{R}\} \\ &\quad \cup \{n(R) \mid R \in K \cap \mathcal{R}\} \\ \mathcal{A}^n &= (\mathcal{A} \setminus \mathcal{R}) \cup \{n(R) \mid R \in \mathcal{R}\} \end{aligned}$$

We define that any abducible literal  $L$  has the name  $L$ , i.e.,  $n(L) = L$ . It is shown in [12], that for any observation  $O$  there is a 1-1 correspondence between (anti-)explanations with respect to  $\langle K, A \rangle$  and those with respect to  $\langle K^n, A^n \rangle$ . That is, for  $n(E) = \{n(R) | R \in E\}$  and  $n(F) = \{n(R) | R \in F\}$ : an observation  $O$  has a (minimal) skeptical (anti-)explanation  $(E, F)$  with respect to  $\langle K, A \rangle$  iff  $O$  has a (minimal) skeptical (anti-)explanation  $(n(E), n(F))$  with respect to  $\langle K^n, A^n \rangle$ . Hence, insertion (deletion) of a rule's name in the normal form corresponds to insertion (deletion) of the rule in the original program. In sum, with the normal form transformation, any abductive program with abducible rules is reduced to an abductive program with only abducible literals.

*Example 5.* We transform the example knowledge base  $K$  into its normal form based on a set of abducibles that is identical to  $K$ : that is  $\mathcal{A} = K$ ; a similar setting will be used in Section 5.2 to achieve deletion of formulas from  $K$ . Hence we transform  $\langle K, \mathcal{A} \rangle$  into its normal form  $\langle K^n, \mathcal{A}^n \rangle$  as follows where we write  $n(R)$  for the naming atom of the only rule in  $\mathcal{A}$ :

$$\begin{aligned} K^n &= \{Ill(\text{Mary}, \text{Aids}), \quad Treat(\text{Pete}, \text{Medi1}), \quad n(R), \\ &\quad Ill(x, \text{Aids}); Ill(x, \text{Flu}) \leftarrow Treat(x, \text{Medi1}), not\ Treat(x, \text{Medi2}), n(R)\} \\ \mathcal{A}^n &= \{Ill(\text{Mary}, \text{Aids}), \quad Treat(\text{Pete}, \text{Medi1}), \quad n(R) \} \end{aligned}$$

## 4.2 Update programs

Minimal (anti-)explanations can be computed with *update programs* (UPs) [12]. The *update-minimal* (U-minimal) answer sets of a UP describe which rules have to be deleted from the program, and which rules have to be inserted into the program, in order (un-)explain an observation.

For the given EDP  $K$  and a given set of abducibles  $\mathcal{A}$ , a set of **update rules**  $UR$  is devised that describe how entries of  $K$  can be changed. This is done with the following three types of rules.

1. **[Abducible rules]** The rules for abducible literals state that an abducible is either true in  $K$  or not. For each  $L \in \mathcal{A}$ , a new atom  $\bar{L}$  is introduced that has the same variables as  $L$ . Then the set of abducible rules for each  $L$  is defined as

$$abd(L) := \{L \leftarrow not\bar{L}, \bar{L} \leftarrow notL\}.$$

2. **[Insertion rules]** Abducible literals that are not contained in  $K$  might be inserted into  $K$  and hence might occur in the set  $E$  of the explanation  $(E, F)$ . For each  $L \in \mathcal{A} \setminus K$ , a new atom  $+L$  is introduced and the insertion rule is defined as

$$+L \leftarrow L.$$

3. **[Deletion rules]** Abducible literals that are contained in  $K$  might be deleted from  $K$  and hence might occur in the set  $F$  of the explanation  $(E, F)$ . For each  $L \in \mathcal{A} \cap K$ , a new atom  $-L$  is introduced and the deletion rule is defined as

$$-L \leftarrow notL.$$

The **update program** is then defined by replacing abducible literals in  $K$  with the update rules; that is,

$$UP = (K \setminus \mathcal{A}) \cup UR.$$

*Example 6.* Continuing Example 5, from  $\langle K^n, \mathcal{A}^n \rangle$  we obtain

$$\begin{aligned} UP = \{ & abd(ill(\text{Mary}, \text{Aids})), \quad abd(\text{Treat}(\text{Pete}, \text{Medi1})), \quad abd(n(R)), \\ & -ill(\text{Mary}, \text{Aids}) \leftarrow not\,ill(\text{Mary}, \text{Aids}), \\ & -\text{Treat}(\text{Pete}, \text{Medi1}) \leftarrow not\,\text{Treat}(\text{Pete}, \text{Medi1}), \\ & -n(R) \leftarrow not\,n(R), \\ & ill(x, \text{Aids}); ill(x, \text{Flu}) \leftarrow \text{Treat}(x, \text{Medi1}), not\,\text{Treat}(x, \text{Medi2}), n(R) \} \end{aligned}$$

The set of atoms  $+L$  is the set  $\mathcal{UA}^+$  of positive update atoms; the set of atoms  $-L$  is the set  $\mathcal{UA}^-$  of negative update atoms. The set of **update atoms** is  $\mathcal{UA} = \mathcal{UA}^+ \cup \mathcal{UA}^-$ . From all answer sets of an update program  $UP$  we can identify those that are **update minimal** (U-minimal): they contain less update atoms than others. Thus,  $S$  is U-minimal iff there is no answer set  $T$  such that  $T \cap \mathcal{UA} \subset S \cap \mathcal{UA}$ .

### 4.3 Ground observations

It is shown in [9] how in some situations the observation formulas  $O$  can be mapped to new positive ground observations. Non-ground atoms with variables can be mapped to a new ground observation. Several positive observations can be conjoined and mapped to a new ground observation. A negative observation (for which an anti-explanation is sought) can be mapped as a NAF-literal to a new positive observation (for which then an explanation has to be found). Moreover, several negative observations can be mapped as a conjunction of NAF-literals to one new positive observation such that its resulting explanation acts as an anti-explanation for all negative observations together. Hence, in extended abduction it is usually assumed that  $O$  is a positive ground observation for which an explanation has to be found. In case of finding a skeptical explanation, an inconsistency check has to be made on the resulting knowledge base. Transformations to a ground observation and inconsistency check will be detailed in Section 5.1 and applied to confidentiality-preservation.

## 5 Confidentiality-Preservation with UPs

We now show how to achieve confidentiality-preservation by extended abduction: we define the set of abducibles and describe how a confidentiality-preserving knowledge base can be obtained by computing U-minimal answer sets of the appropriate update program. We additionally distinguish between the case that we allow only deletions of formulas – that is, in the anti-explanation  $(E, F)$  the set  $E$  of positive anti-explanation formulas is empty – and the case that we also allow insertions.

### 5.1 Policy transformation for credulous users

Elements of the confidentiality policy will be treated as negative observations for which an anti-explanation has to be found. Accordingly, we will transform policy elements to a set of rules containing new positive observations as sketched in Section 4.3. We will call these rules **policy transformation rules for credulous users** ( $PTR^{cred}$ ).

More formally, assume *policy* contains  $k$  elements. For each conjunction  $C_i \in \text{policy}$  ( $i = 1 \dots k$ ), we introduce a new negative ground observation  $O_i^-$  and map  $C_i$  to  $O_i^-$ . As each



$C_i$  is a conjunction of (NAF-)literals, the resulting formula is an EDP rule. As a last policy transformation rule, we add one that maps all new negative ground observations  $O_i^-$  (in their NAF version) to a positive observation  $O^+$ . Hence,

$$PTR^{cred} := \{O_i^- \leftarrow C_i \mid C_i \in policy\} \cup \{O^+ \leftarrow not\ O_1^-, \dots, not\ O_k^-\}.$$

*Example 7.* The set of policy transformation rules for  $policy'$  is

$$PTR^{cred} = \{O_1^- \leftarrow Ill(x, Aids) \ , \ O_2^- \leftarrow \neg AbleToWork(x) \ , \ O^+ \leftarrow not\ O_1^-, not\ O_2^-\}$$

Lastly, we consider a **goal rule**  $GR$  that enforces the single positive observation  $O^+$ :  $GR = \{\leftarrow not\ O^+\}$ .

We can also allow more expressive policy elements in disjunctive normal form (DNF: a disjunction of conjunctions of (NAF-)literals). If we map a DNF formula to a new observation (that is,  $O_{disj}^- \leftarrow C_1 \vee \dots \vee C_l$ ) this is equivalent to mapping each conjunct to the observation (that is,  $O_{disj}^- \leftarrow C_1, \dots, O_{disj}^- \leftarrow C_l$ ). We also semantically justify this splitting into disjuncts by arguing that in order to protect confidentiality of a disjunctive formula we indeed have to protect each disjunct alone. However, if variables are shared among disjuncts, these variables have to be grounded according to the Herbrand universe of  $K \cup prior$  first; otherwise the shared semantics of these variables is lost.

## 5.2 Deletions for credulous users

As a simplified setting, we first of all assume that only deletions are allowed to achieve confidentiality-preservation. This setting can informally be described as follows: For a given knowledge base  $K$ , if we only allow deletions of rules from  $K$ , we have to find a *skeptical negative explanation*  $F$  that explains the new positive observation  $O^+$  while respecting  $prior$  as invariable a priori knowledge. The set of abducibles is thus identical to  $K$  as we want to choose formulas from  $K$  for deletion:  $\mathcal{A} = K$ . That is, in total we consider the abductive program  $\langle K, \mathcal{A} \rangle$ . Then, we transform it into normal form  $\langle K^n, \mathcal{A}^n \rangle$ , and compute its update program  $UP$  as described in Section 4.2. As for  $prior$ , we add this set to the update program  $UP$  in order to make sure that the resulting answer sets of the update program do not contradict  $prior$ . Finally, we add all the policy transformation rules  $PTR^{cred}$  and the goal rule  $GR$ . The goal rule is then meant as a constraint that filters out those answer sets of  $UP \cup prior \cup PTR^{cred}$  in which  $O^+$  is *true*. We thus obtain a new program  $P$  as

$$P = UP \cup prior \cup PTR^{cred} \cup GR$$

and compute its U-minimal answer sets. If  $S$  is one of these answer sets, the negative explanation  $F$  is obtained from the negative update atoms contained in  $S$ :  $F = \{L \mid -L \in S\}$ .

To obtain a confidentiality-preserving knowledge base for a credulous user, we have to check for inconsistency with the negation of the positive observation  $O^+$  (which makes  $F$  a *skeptical* explanation of  $O^+$ ); and allow only answer sets of  $P$  that are U-minimal among those respecting this inconsistency property. More precisely, we check whether

$$(K \setminus F) \cup prior \cup PTR^{cred} \cup \{\leftarrow O^+\} \text{ is inconsistent.} \quad (1)$$

*Example 8.* We combine the update program  $UP$  of  $K$  with  $prior$  and the policy transformation rules and goal rule. This leads to the following two U-minimal answer sets with only deletions which satisfy the inconsistency property (1):

$$\begin{aligned} S'_1 &= \{-Ill(\text{Mary}, \text{Aids}), -Treat(\text{Pete}, \text{Medi1}), n(R), \overline{Ill(\text{Mary}, \text{Aids})}, \overline{Treat(\text{Pete}, \text{Medi1})}, O^+\} \\ S'_2 &= \{-Ill(\text{Mary}, \text{Aids}), Treat(\text{Pete}, \text{Medi1}), -n(R), \overline{Ill(\text{Mary}, \text{Aids})}, \overline{n(R)}, O^+\}. \end{aligned}$$

These answer sets correspond to the minimal solutions from Example 4 where  $Ill(\text{Mary}, \text{Aids})$  must be deleted together with either  $Treat(\text{Pete}, \text{Medi1})$  or the rule named  $R$ .

**Theorem 1 (Correctness for deletions).** *A knowledge base  $K^{pub} = K \setminus F$  preserves confidentiality and changes  $K$  subset-minimally iff  $F$  is obtained by an answer set of the program  $P$  that is U-minimal among those satisfying the inconsistency property (1).*

*Proof. (Sketch)* First of all note that because we chose  $K$  to be the set of abducibles  $\mathcal{A}$ , only negative update atoms from  $\mathcal{UA}^-$  occur in  $UP$  – no insertions with update atoms from  $\mathcal{UA}^+$  will be possible. Hence we automatically obtain an anti-explanation  $(E, F)$  where  $E$  is empty. As shown in [12], there is a 1-1 correspondence of minimal explanations and U-minimal answer sets of update programs; and anti-explanations are identical to explanations of a new positive observation when applying the transformations as in  $PTR^{cred}$ . By properties of skeptical (anti-)explanations we have thus  $K^{pub} \cup prior \cup PTR^{cred} \models O^+$  but for every  $O_i^-$  there is no answer set in which  $O_i^-$  is satisfied. This holds iff for every policy element  $C_i$  there is no answer set of  $K^{pub} \cup prior$  that satisfies any instantiation of  $C_i$  (with respect to the Herbrand universe of  $K^{pub} \cup prior$ ); thus  $cred(K^{pub} \cup prior, C_i) = \emptyset$ . Subset-minimal change carries over from U-minimality of answer sets.

### 5.3 Deletions and literal insertions

To obtain a confidentiality-preserving knowledge base, (incorrect) entries may also be inserted into the knowledge base. To allow for insertions of literals, a more complex set  $\mathcal{A}$  of abducibles has to be chosen. We reinforce the point that the subset  $\mathcal{A} \cap K$  of abducibles that are already contained in the knowledge base  $K$  are those that may be deleted while the subset  $\mathcal{A} \setminus K$  of those abducibles that are not contained in  $K$  may be inserted.

First of all, we assume that the policy transformation is applied as described in Section 5.1. Then, starting from the new negative observations  $O_i^-$  used in the policy transformation rules, we trace back all rules in  $K \cup prior \cup PTR^{cred}$  that influence these new observations and collect all literals in the bodies of these rules. In other words, we construct a dependency graph (as in [16]) and collect the literals that the negative observations depend on. More formally, let  $P_0$  be the set of literals that the new observations  $O_i^-$  directly depend on:

$$\begin{aligned} P_0 &= \{L \mid L \in body(R) \text{ or } notL \in body(R) \\ &\quad \text{where } R \in PTR^{cred} \text{ and } O_i^- \in head(R)\} \end{aligned}$$

Next we iterate and collect all the literals that the  $P_0$  literals depend on:

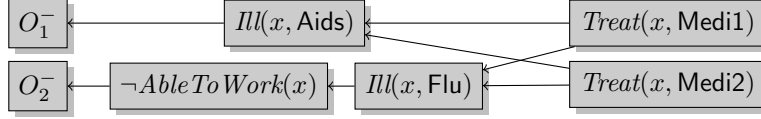
$$\begin{aligned} P_{j+1} &= \{L \mid L \in body(R) \text{ or } notL \in body(R) \\ &\quad \text{where } R \in K \cup prior \cup PTR^{cred} \text{ and } head(R) \cap P_j \neq \emptyset\} \end{aligned}$$

and combine all such literals in a set  $\mathcal{P} = \bigcup_{j=0}^{\infty} P_j$ .

As we also want to have the option to delete rules from  $K$  (not only the literals in  $\mathcal{P}$ ), we define the set of abducibles as the set  $\mathcal{P}$  plus all those rules in  $K$  whose head depends on literals in  $\mathcal{P}$ :

$$\mathcal{A} = \mathcal{P} \cup \{R \mid R \in K \text{ and } \text{head}(R) \cap \mathcal{P} \neq \emptyset\}$$

*Example 9.* For the example  $K \cup \text{prior} \cup \text{PTR}^{\text{cred}}$ , the dependency graph is shown in Figure 2. We note that the new negative observation  $O_1^-$  directly depends on the literal  $\text{Ill}(x, \text{Aids})$  and the new negative observation  $O_2^-$  directly depends on the literal  $\neg \text{AbleToWork}(x)$ ; this is the first set of literals  $P_0 = \{\text{Ill}(x, \text{Aids}), \neg \text{AbleToWork}(x)\}$ . By tracing back the dependencies in the graph,  $\mathcal{P} = \{\text{Ill}(x, \text{Aids}), \neg \text{AbleToWork}(x), \text{Ill}(x, \text{Flu}), \text{Treat}(x, \text{Medi1}), \text{Treat}(x, \text{Medi2})\}$  is obtained. Lastly, we also have to add the rule  $R$  from  $K$  to  $\mathcal{A}$  because literals in its head are contained in  $\mathcal{P}$ .



**Fig. 2.** Dependency graph for literals in  $K \cup \text{prior} \cup \text{PTR}$

We obtain the normal form and then the update program  $UP$  for  $K$  and the new set of abducibles  $\mathcal{A}$ . The process of finding a skeptical explanation proceeds with finding an answer set of program  $P$  as in Section 5.2 where additionally the positive explanation  $E$  is obtained as  $E = \{L \mid +L \in S\}$  and  $S$  is U-minimal among those satisfying

$$(K \setminus F) \cup E \cup \text{prior} \cup \text{PTR}^{\text{cred}} \cup \{\leftarrow O^+\} \text{ is inconsistent.} \quad (2)$$

*Example 10.* For  $UP$  from Example 8 the new set of abducibles leads to additional insertion rules. Among others, the insertion rule for the new abducible  $\text{Treat}(\text{Pete}, \text{Medi2})$  is  $+\text{Treat}(\text{Pete}, \text{Medi2}) \leftarrow \text{Treat}(\text{Pete}, \text{Medi2})$ . With this new rule included in  $UP$ , we also obtain the solution of Example 4 where the fact  $\text{Treat}(\text{Pete}, \text{Medi2})$  is inserted into  $K$  (together with deletion of  $\text{Ill}(\text{Mary}, \text{Aids})$ ).

**Theorem 2 (Correctness for deletions & literal insertions).** *A knowledge base  $K^{\text{pub}} = (K \setminus F) \cup E$  preserves confidentiality and changes  $K$  subset-minimally iff  $(E, F)$  is obtained by an answer set of program  $P$  that is U-minimal among those satisfying inconsistency property (2).*

*Proof. (Sketch)* In  $UP$ , positive update atoms from  $\mathcal{UA}^+$  occur for literals on which the negative observations depend. For subset-minimal change, only these literals are relevant for insertions; inserting other literals will lead to non-minimal change. In analogy to Theorem 1, by the properties of minimal skeptical (anti-)explanations that correspond to U-minimal answer sets of an update program, we obtain a confidentiality-preserving  $K^{\text{pub}}$  with minimal change.

## 6 Discussion and Conclusion

This article showed that when publishing a logic program, confidentiality-preservation can be ensured by extended abduction; more precisely, we showed that under the credulous query response it reduces to finding skeptical anti-explanations with update programs. This is an application of data modification, because a user can be misled by the published knowledge base to believe incorrect information; we hence apply dishonesties [11] as a security mechanism. This is in contrast to [16] whose aim is to avoid incorrect deductions while enforcing access control on a knowledge base. Another difference to [16] is that they do not allow disjunctions in rule heads; hence, to the best of our knowledge this article is the first one to handle a confidentiality problem for EDPs. In [3] the authors study databases that may provide users with incorrect answers to preserve security in a multi-user environment. Different from our approach, they consider a database as a set of formulas of propositional logic and formulate the problem using modal logic. In analogy to [12], a complexity analysis for our approach can be achieved by reduction of extended abduction to normal abduction. Work in progress covers data publishing for skeptical users; future work might handle insertion of non-literal rules.

## References

1. Foto N. Afrati and Phokion G. Kolaitis. Repair checking in inconsistent databases: algorithms and complexity. In *ICDT2009*, volume 361 of *ACM International Conference Proceeding Series*, pages 31–41. ACM, 2009.
2. Joachim Biskup. Usability confinement of server reactions: Maintaining inference-proof client views by controlled interaction execution. In *DNIS 2010*, volume 5999 of *LNCS*, pages 80–106. Springer, 2010.
3. Piero A. Bonatti, Sarit Kraus, and V. S. Subrahmanian. Foundations of secure deductive databases. *IEEE Trans. Knowl. Data Eng.*, 7(3):406–422, 1995.
4. Francesco Calimeri, Giovambattista Ianni, Francesco Ricca, Mario Alviano, Annamaria Bria, Gelsomina Catalano, Susanna Cozza, Wolfgang Faber, Onofrio Febraro, Nicola Leone, Marco Manna, Alessandra Martello, Claudio Panetta, Simona Perri, Kristian Reale, Maria Carmela Santoro, Marco Sirianni, Giorgio Terracina, and Pierfrancesco Veltri. The third answer set programming competition: Preliminary report of the system competition track. In *LPNMR 2011*, volume 6645 of *LNCS*, pages 388–403. Springer, 2011.
5. Jürgen Dix, Wolfgang Faber, and V. S. Subrahmanian. The relationship between reasoning about privacy and default logics. In *LPAR 2005*, volume 3835 of *Lecture Notes in Computer Science*, pages 637–650. Springer, 2005.
6. Csilla Farkas and Sushil Jajodia. The inference problem: A survey. *SIGKDD Explorations*, 4(2):6–11, 2002.
7. Michael Gelfond and Vladimir Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386, 1991.
8. Bernardo Cuenca Grau and Ian Horrocks. Privacy-preserving query answering in logic-based information systems. In *ECAI2008*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 40–44. IOS Press, 2008.
9. Katsumi Inoue and Chiaki Sakama. Abductive framework for nonmonotonic theory change. In *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI 95)*, volume 1, pages 204–210. Morgan Kaufmann, 1995.
10. Jiefei Ma, Alessandra Russo, Krysia Broda, and Emil Lupu. Multi-agent confidential abductive reasoning. In *ICLP (Technical Communications)*, volume 11 of *LIPICs*, pages 175–186. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2011.
11. Chiaki Sakama. Dishonest reasoning by abduction. In *22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 1063–1064. IJCAI/AAAI, 2011.
12. Chiaki Sakama and Katsumi Inoue. An abductive framework for computing knowledge base updates. *Theory and Practice of Logic Programming*, 3(6):671–713, 2003.
13. Phiniki Stouppa and Thomas Studer. Data privacy for knowledge bases. In Sergei N. Artëmov and Anil Nerode, editors, *LFCS2009*, volume 5407 of *LNCS*, pages 409–421. Springer, 2009.

14. Tyrone S. Toland, Csilla Farkas, and Caroline M. Eastman. The inference problem: Maintaining maximal availability in the presence of database updates. *Computers & Security*, 29(1):88–103, 2010.
15. Lena Wiese. Horizontal fragmentation for data outsourcing with formula-based confidentiality constraints. In *IWSEC 2010*, volume 6434 of *LNCS*, pages 101–116. Springer, 2010.
16. Lingzhong Zhao, Junyan Qian, Liang Chang, and Guoyong Cai. Using ASP for knowledge management with user authorization. *Data & Knowl. Eng.*, 69(8):737–762, 2010.